

IN THE UNITED STATES PATENT & TRADEMARK OFFICE

IN RE APPLICATION OF: Nam, et al.

SERIAL NO.:

FILED:

**TITLE: PREPROCESSING OF DIGITAL AUDIO DATA FOR IMPROVING
 PERCEPTUAL SOUND QUALITY ON A MOBILE PHONE**

DKT. NO.: KIM-10188

CLAIM FOR PRIORITY

Commissioner For Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

The benefit of the filing date of the following prior foreign application filed in the following foreign country is hereby requested for the above-identified application and the priority provided in 35 U.S.C. §119 is hereby claimed:

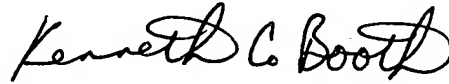
Korean Patent Application No. 10-2003-0001330

In support of this claim, a certified copy of the original foreign application:

- ☒ is filed herewith.
- ☐ was filed on _____ in parent application Serial No. _____, filed _____, 19__.

It is requested that the file of this application be marked to indicate that the requirements of 35 U.S.C. §119 have been fulfilled and that the Patent and Trademark Office kindly acknowledge receipt of this document.

Respectfully submitted,

A handwritten signature in black ink, reading "Kenneth C. Booth". The signature is written in a cursive, flowing style.

Date: January 8, 2004

Kenneth C. Booth
Reg. No. 42,342

SCHMEISER, OLSEN & WATTS LLP
18 E. University Dr., #101
Mesa, Arizona 85201
(480) 655-0073



별첨 사본은 아래 출원의 원본과 동일함을 증명함.

This is to certify that the following application annexed hereto is a true copy from the records of the Korean Intellectual Property Office.

출원 번호 : 10-2003-0001330
Application Number

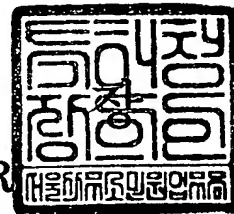
출원 년 월 일 : 2003년 01월 09일
Date of Application JAN 09, 2003

출원인 : 와이더텐닷컴 주식회사
Applicant(s) WIDER THAN.COM CO., LTD.



2003 년 11 월 28 일

특 허 청
COMMISSIONER



【서지사항】

【서류명】	특허출원서
【권리구분】	특허
【수신처】	특허청장
【참조번호】	4983
【제출일자】	2002.01.09
【발명의 명칭】	이동 전화상의 체감 음질을 향상시키기 위한 디지털 오디오 신호의 전처리 방법
【발명의 영문명칭】	Preprocessing of Digital Audio data for Improving Perceptual Sound Quality on a Mobile Phone
【출원인】	
【명칭】	와이더덴닷컴 주식회사
【출원인코드】	1-2000-053865-3
【대리인】	
【성명】	주성민
【대리인코드】	9-1998-000517-7
【대리인】	
【성명】	백만기
【대리인코드】	9-1999-000500-0
【발명자】	
【성명의 국문표기】	남영한
【성명의 영문표기】	NAM, Young Han
【주민등록번호】	760223-1106427
【우편번호】	121-764
【주소】	서울특별시 마포구 대흥동 태영아파트 107-1204
【국적】	KR
【발명자】	
【성명의 국문표기】	박섭형
【성명의 영문표기】	PARK, Seop Hyeong
【주민등록번호】	610921-1056011
【우편번호】	156-821
【주소】	서울특별시 동작구 사당4동 278-119
【국적】	KR

【발명자】

【성명】

전윤희

【출원인코드】

4-1998-715457-4

【취지】

특허법 제42조의 규정에 의하여 위와 같이 출원합니다. 대리인
주성민 (인) 대리인
백만기 (인)

【수수료】

【기본출원료】

20 면 29,000 원

【가산출원료】

29 면 29,000 원

【우선권주장료】

0 건 0 원

【심사청구료】

0 항 0 원

【합계】

58,000 원

【감면사유】

중소기업

【감면후 수수료】

29,000 원

【첨부서류】

1. 요약서·명세서(도면)_1통 2. 위임장_1통 3. 중소기업기본법시행령 제2조에 의한 중소기업에 해당함을 증명하는 서류[사업자등록증 사본]_1통 4. 중소기업기본법시행령 제2조에 의한 중소기업에 해당함을 증명하는 서류[원천징수이행상황신고서 사본]_1통

【요약서】**【요약】**

최근 휴대전화의 사용 영역이 점차 넓어지고 휴대전화로 음악을 듣는 서비스가 활성화됨에 따라 휴대전화를 통해 듣는 음악의 음질 개선이 널리 요구되고 있다. 휴대전화는 EVRC 및 그와 상응하는 음성 부호화 방법에 의해 음성에 최적화된 방법으로 음성을 부호화하는데, 이와 같이 음성에 최적화된 부호화 기법에 의해 부호화된 음악 신호는 왜곡이 심하며, 배경 잡음이 없더라도 음성 부호화 방법 자체의 특성에 따른 끊김 현상이 광범위하게 발생하고 있다.

본 발명은 휴대전화에서 음악을 들을 경우 음성 부호화 자체의 특성에 따라 끊김 현상이 발생하는 것을 방지하는 전처리 방법을 제안하는 것을 그 목적으로 하며, 나아가, 휴대전화에서 듣는 음악 음질의 최적화를 추구하였다.

이를 위해 피치를 강화하여 EVRC의 전송률 결정 알고리즘에서 장기 예측이득을 증가시켜 음악 신호가 높은 전송률로 전송될 수 있도록 하는 한편, 이동 통신 단말기에서의 음질을 최적화하기 위하여 음폭 압축을 수행함으로써 큰 소리의 포화 현상과 작은 소리의 끊어짐 현상을 방지할 수 있었다.

【대표도】

도 6

【색인어】

음폭 압축, 피치 강화, 심리 음향 모델, AGC, 다중 톤 노치 필터, 오차 침두치 강화

【명세서】

【발명의 명칭】

이동 전화상의 체감 음질을 향상시키기 위한 디지털 오디오 신호의 전처리 방법
{Preprocessing of Digital Audio data for Improving Perceptual Sound Quality on a Mobile Phone}

【도면의 간단한 설명】

도 1은 EVRC(Enhanced Variable Rate Coding) 부호화기의 고수준 블록도.

도 2a는 통화 음성을 EVRC로 부호화할 경우 추정 잡음의 변화 추이를 나타내는 그래프.

도 2b는 음악을 EVRC로 부호화할 경우 추정 잡음의 변화 추이를 나타내는 그래프.

도 3a는 통화 음성을 EVRC로 부호화할 경우 전송률 결정 임계치(RDT: Rate Determination Threshold)의 변화 추이를 나타내는 그래프.

도 3b는 음악을 EVRC로 부호화할 경우 전송률 결정 임계치의 변화 추이를 나타내는 그래프.

도 4a는 주 피치 성분이 존재하는 음악 신호에 대한 프레임 예측 오차 신호를 나타내는 그래프.

도 4b는 주 피치 성분이 존재하지 않는 음악 신호에 대한 프레임 예측 오차 신호를 나타내는 그래프.

도 5a는 주 피치 성분이 존재하는 음악 신호에 대한 프레임 오차 자기상관을 나타내는 그래프.

도 5b는 주 피치 성분이 존재하지 않는 음악 신호에 대한 프레임 오차 자기상관을 나타내는 그래프.

도 6은 본 발명에 따른 음악 전처리 과정을 나타내는 개념도.

도 7은 본 발명에 따른 AGC 전처리를 구현하기 위한 플로우 차트.

도 8은 Terhardt의 최대 가청 임계치(ATH: Absolute Threshold of Hearing) 모델을 나타내는 그래프.

도 9는 임계 대역폭(Critical Bandwidth)을 설명하기 위한 그래프.

도 10은 본 발명에 따른 피치 강화를 구현하기 위한 블록도.

도 11은 본 발명에 따른 다중 톤 노치 필터(MTNF: Multi-Tone Notch Filter)를 적용시켰을 경우의 스펙트럼 변화를 나타내는 그래프.

도 12는 본 발명에 따른 다중 톤 노치 필터를 적용시켰을 경우의 $R_{\varepsilon}[0]$ 의 변화를 나타내는 그래프.

도 13은 본 발명에 따른 전처리 방법을 적용시켰을 경우의 대역 에너지와 전송률 결정 임계치의 변화를 나타내는 그래프.

도 14는 본 발명에 따른 전처리 방법을 적용시켰을 경우의 전처리 결과 신호를 나타내는 그래프.

【발명의 상세한 설명】

【발명의 목적】

【발명이 속하는 기술분야 및 그 분야의 종래기술】

- <19> 본 발명은 이동 전화에서 사용되는 음성 코덱을 통해서 오디오 파일을 압축/복원할 때 발생하는, 수신단에서 음악의 음질 저하를 방지하기 위하여 음악 파일을 전처리하는 방법에 관한 것이다.
- <20> 이동 통신 시스템의 음성 채널 대역폭은 유선 통신 시스템의 64 kbps에 비해서 매우 작기 때문에, 음성 신호는 압축하여 전송된다. 현재 이동 통신 시스템에서 사용되는 음성 압축 기법으로는 IS-95의 QCELP(Qualcomm Code Excited Linear Prediction), EVRC(Enhanced Variable Rate Coding), GSM의 VSELP(Vector-Sum Excited Linear Prediction), PRE-LTP(Regular-Pulse Excited LPC with a Long-Term Predictor), ACELP(Algebraic Code Excited Linear Prediction) 등이 있는데, 모두 LPC(Linear Prediction Coding: 선형 예측 부호화) 분석 방법을 기반으로 하는 공통점을 가지고 있다. LPC 계열의 음성 압축 기법은 사람의 발성 구조에 최적화된 모델을 사용하고 있어서 사람의 음성을 중전송률이나 저전송률로 압축하는데 매우 효율적이다. 또한, 스펙트럼의 효율을 높이고 시스템의 소비 전력을 줄이기 위해서 사람이 말을 할 때만 신호를 압축하여 전달하고 사람이 말을 하지 않을 때는 신호를 전달하지 않는 방법을 사용하는데, 이 때 음성 부분을 검출하는 기능을 VAD(voice activity detection: 음성 활동 검출)라 한다.
- <21> 최근에 이동전화 가입자에게 음악을 제공하는 서비스, 수신대기 상태에서 신호음 대신에 가입자가 지정한 오디오 파일을 재생해 주는 서비스("컬러링" 서비스) 등이 널리 쓰임에 따라

이동 통신 단말기에 음악을 전송하는 경우가 점점 늘고 있다. 그러나, LPC 계열의 음성 압축 기법은 사람의 음성이 아닌, 음악이나 기타 주파수 스펙트럼이 복잡한 신호를 압축하기에는 적당하지 않다. 따라서, 일반적으로 가청 주파수 전대역을 포함하는 오디오 신호를 이동 통신 단말기를 통해서 들을 경우에 소리가 일시적으로 단절되거나 일부 대역의 소리만 들리게 되는 등 신호의 왜곡이 심하게 발생하며, 이러한 현상이 이동통신 시스템에서 음악 서비스를 제공하는데 가장 큰 장애 요인이 되고 있다.

<22> 본 발명은 이동 통신 시스템의 단말기에서 청취하는 오디오 신호의 품질을 향상시키기 위하여, 전송하려는 원천 음악 파일을 전처리하는 방법을 제공한다. 구체적으로는, 오디오 신호를 EVRC 코덱을 통과시킨 후에 발생하는 다양한 문제점을 파악하여 이를 완화할 수 있는 전처리 방법을 제공한다. 본 명세서에서 제안하는 전처리 방법은 기존의 이동 통신 시스템 자체에 변형을 가하지 않으면서도 수신 음악의 품질을 향상시킬 수 있는 방법이기 때문에 현실적으로 매우 유용하다. 이러한 접근 방법은 EVRC 코덱 이외의 다른 코덱에서도 유사하게 적용될 수 있다.

【발명이 이루고자 하는 기술적 과제】

<23> 음악 신호의 전송률을 높여 음악의 끊김 현상이 발생하는 것을 현저하게 줄임으로써, 궁극적으로는 휴대전화에서 듣는 음악 음질을 최적화하는 것을 그 목적으로 한다.

【발명의 구성 및 작용】

<24> 먼저, 음성 코덱, 특히 EVRC 코덱으로 오디오 신호를 압축할 때 음질이 저하되는 현상과 그 원인을 살펴본다. 코덱에 따라 조금씩 다르기는 하지만 이러한 문제점들은 LPC 계열의 코

텍에서 공통적으로 발생한다. 음질이 저하되는 현상은 다음과 같은 다섯 가지로 분류해 볼 수 있다.

- <25> ① 고주파 대역 성분 완전 손실 현상
- <26> ② 저주파 대역의 일부 주파수 성분 손실 현상
- <27> ③ 페이드 인/페이드 아웃 부분의 재생 불량 현상
- <28> ④ 음악 소리가 갑자기 작아지는 현상
- <29> ⑤ 음악 중간 부분의 끊김 현상

<30> 먼저, 첫 번째 현상의 원인을 살펴보면, 일반적으로 오디오 신호는 가청주파수인 20 ~ 20,000 Hz 사이의 신호들을 모두 포함하고 있으나, 대부분의 협대역 음성 코덱은 신호를 압축하기 전에 4 KHz(또는 3.4 KHz) 이하의 저역통과 필터 (lowpass filter)를 사용하여 고주파 성분을 제거하기 때문에, 원래의 음악에 포함되어 있던 고주파 성분이 완전히 손실되게 된다. 이러한 문제점은 음악을 압축하는 과정에서 수반되는 일반적으로 피할 수 없는 부분이다.

<31> 두 번째 현상이 발생하는 것은 LPC 계열의 음성 압축 기법의 근본적인 문제점 때문이다. 일반적인 LPC 계열의 압축 기법은 입력 신호의 피치와 포먼트 주파수(formant frequency)를 구하여, 이 둘을 이용하여 합성한 신호와 입력 신호의 차이를 최소화하는 여기 신호(excitation signal)를 정해진 부호책(codebook)에서 구한다. 사람의 음성과 달리 다양한 화음이 혼재하는 음악에서는 피치를 명확하게 구분하는 것이 쉽지 않으며, 사람의 음성과 음악의 포먼트 성분도 확연한 차이가 있다. 결과적으로, 예측 오차 신호를 최소화하는 과정에서 오디오 신호는 음성 신호에 비해서 오차가 매우 클 수 밖에 없으며, 이 과정에서 원 오디오 신호에 포함되어 있는 다양한 주파수 성분들이 충실하게 표현되지 못하고 손실되는 경우가 많다.

- <32> 세 번째의 페이드 인, 페이드 아웃 부분이 재생되지 않는 현상은 음성 코덱이 일정 크기 이하의 소리를 잡음으로 처리하기 때문에 발생된다.
- <33> 네 번째로, 갑자기 음악 소리가 작아지는 현상은 주로 음악 소리의 전력의 변화율이 일정 시간 이상 비슷한 값으로 유지되면 EVRC 부호기의 전처리단에서 배경 잡음 레벨을 증가시키면서 입력 신호에 1보다 작은 이득(gain)을 곱하므로, 상대적으로 큰 소리도 갑자기 작아지기 때문에 나타난다.
- <34> 마지막으로 다섯 번째인 음악의 중간 부분이 끊기는 현상의 원인은 EVRC의 가변 전송률 때문이다. 본 명세서에서는 이러한 음악의 중간 부분이 끊기는 현상을 해결하는 방법을 제공한다. EVRC 부호기는 음성 신호를 1, 1/2, 1/8의 세 가지 전송률 유형으로 구분하여 달리 처리한다. 이 가운데 1/8 전송률은 EVRC 부호기가 입력 신호를 음성이 아닌 잡음이라고 판단한 것을 의미한다. 드럼과 같은 타악기 소리는 음성 코덱이 잡음으로 처리하기 쉬운 스펙트럼 성분을 가지고 있어서 이러한 소리가 많이 포함된 음악에서는 끊김 현상이 자주 발생한다. 또한, 사람 목소리와는 달리 음악은 소리의 크기가 자주 변할 수 있기 때문에 에너지 레벨이 낮은 부분에서도 끊김 현상이 발생한다. 이 현상은 VAD에 근거한 DTX(discontinuous transmission)를 채택한 모든 시스템에서 공통적으로 나타난다. 이 문제점은 EVRC 코덱의 전송률 결정 과정에서 오디오 신호의 모든 프레임이 전송률 1로 부호화되도록 미리 전처리를 함으로써 해결할 수 있다. 비록 본 명세서에서는 EVRC에 대한 해결 방안을 제시하였으나, 아래의 방법이 다른 가변 전송률 압축 코덱 방법에 의해서도 적용될 수 있다는 것을 당업자라면 이해할 것이다.
- <35> 도 1을 참조하여, EVRC의 전송률 결정 알고리즘(RDA; Rate Decision Algorithm)에 대해 간단히 살펴보기로 한다.

<36> 도 1은 EVRC의 고수준 블록도(high-level block diagram)를 나타낸 것이다. 도 1에서 입력은 8k, 16bit PCM(pulse code modulation) 오디오 신호이고, 부호화된 출력은 전송률 결정 알고리즘(RDA)에 의해 결정된 전송률에 따라 프레임당 171(전송률 1), 80(전송률 1/2), 16(전송률 1/8), 0(블랭크) 비트의 크기를 갖는 디지털 데이터이다. 8k, 16bit PCM 오디오 신호는 160 샘플(20ms) 단위의 프레임으로 분할되어 EVRC 부호화기(encoder; 100)에 입력된다. 잡음 억제 블록(Noise suppression block; 110)은 부호화기의 시작단에 위치하며, 입력된 프레임 신호 $s[n]$ 을 검사하여 잡음이라고 판단되는 프레임에 1보다 작은 이득을 곱하여 해당 프레임 신호를 억제한다. 이 블록을 통과한 신호 $s'[n]$ 은 RDA 블록(120)의 입력으로 사용되며, RDA 블록(120)은 전송률 1, 전송률 1/2, 전송률 1/8, 블랭크 중 하나의 전송률을 결정한다. 부호화 블록(Encoding block; 130)은 이전 단계에서 결정된 전송률에 따라 적절히 파라미터(parameter)를 추출하며, 비트 패킹 블록(Bit packing block; 140)은 추출된 파라미터들을 출력 포맷에 맞도록 패킹(packing)하는 역할을 한다.

<37> 다음 표에 나타난 바와 같이, 최종 부호화된 출력(encoded output)은 RDA에 의해 결정된 전송률에 따라 프레임당 171, 80, 16, 0 비트의 크기를 갖는다.

<38> 【표 1】

프레임 타입	프레임 당 비트
전송률 1	171
전송률 1/2	80
전송률 1/8	16
블랭크	0

<39> RDA 블록(120)은 $s'[n]$ 을 대역통과필터(bandpass filter)를 사용하여 0.3-2.0kHz ($f(1)$)와 2.0-4.0kHz($f(2)$)의 두 개의 대역으로 분리하고, 각 대역 신호 성분($f(1)$, $f(2)$)의 대역 에너지 추정을 배경 잡음(background noise estimate)에 의해 결정되는 전송률 결정 임계치(Rate

Decision Threshold)들과 비교함으로써 전송률을 결정한다. 아래의 식은 상기 두 대역에 대한 각각의 임계치를 계산하는 식이다.

<40>
$$T_{1,i} = k_1(\text{SNR}_{m-1,i})B_{m-1,i}$$

【수학식 1】
$$T_{2,i} = k_2(\text{SNR}_{m-1,i})B_{m-1,i}$$

<41> 여기서, k_1 과 k_2 는 임계치 배율(threshold scale factor)로서 SNR(signal-to-noise ratio) 값이 클수록 큰 값을 갖는 SNR 값에 대한 증가함수이다. 또한, $B_{f(i)}(m-1)$ 는 $m-1$ 번째 프레임의 i 번째 주파수 대역의 추정 잡음이다.

<42> 위 식에 나타난 바와 같이, 전송률 결정 임계치는 추정 잡음에 스케일 계수를 곱한 것으로, 추정 잡음의 크기에 직접 비례한다.

<43> 한편, 대역 에너지는 다음 식에 나타나 바와 같이 각 주파수 대역에 국한된 음성 신호의 자기상관(autocorrelation) 계수의 0에서 16번째 값에 의해 결정된다.

<44>
$$BE_i = R_w[0]R_i[0] + 2.0 \sum_{k=1}^{16} R_w[k]R_i[k]$$

【수학식 2】

<45> 여기서, $BE_{f(i)}$ 는 i 번째 주파수 대역($i=1,2$)의 대역 에너지이고, $R_w(k)$ 는 입력 오디오 신호의 자기상관계수와 관련된 함수이며, $R_{f(i)}(k)$ 는 대역통과 필터의 임펄스 응답(impulse response)의 자기상관계수이다.

<46> 다음으로, 추정 잡음($B_{m,i}$)이 어떻게 갱신되는지 살펴본다. 다음 수학식 3은 추정 잡음의 갱신 과정을 나타내고 있다.

```

<47>      if (  $\beta < 0.30$  for 8 or more consecutive frames )
           $B_{m,i} = \min\{ E^{sm}_{m,i}, 80954304, \max\{1.03B_{m-1,i}, B_{m-1,i} + 1\} \}$ 
        else {
          if (  $SNR_{m-1,i} > 3$  )
             $B_{m,i} = \min\{ E^{sm}_{m,i}, 80954304, \max\{1.00547B_{m-1,i}, B_{m-1,i} + 1\} \}$ 
          else
             $B_{m,i} = \min\{ E^{sm}_{m,i}, 80954304, B_{m-1,i} \}$ 
          }
        if (  $B_{m,i} < lownoise(i)$  )
           $B_{m,i} = lownoise(i)$ 
    【수학식 3】       $m = m + 1$ 

```

<48> 위 수학식에 따르면, m 번째 프레임의 i 번째 주파수 대역의 추정 잡음 $B_{m,i}$ 은 m-1 번째 프레임의 i 번째 주파수 대역의 추정 잡음 $B_{m-1,i}$ 과 m 번째 프레임의 i 번째 주파수 대역의 평활화 대역 에너지(smoothed band energy) $E^{SM}_{m,i}$, 및 m-1 번째 프레임의 i 번째 주파수 대역의 신호대잡음비 $SNR_{m-1,i}$ 에 의해 정해진다.

<49> 위의 수학식 3에서 알 수 있듯이, 8개 이상의 프레임 동안 장기 예측 이득(long-term prediction gain)인 β (계산방법은 후술함)의 값이 0.30보다 작을 경우, 추정 잡음은 평활화 대역 에너지, 이전 프레임의 추정 잡음의 1.03배 및 미리 설정된 배경 추정 잡음의 최대값(80954304) 중 최소값을 취한다. β 의 값이 0.30보다 작지 않을 경우, 추정 잡음은 이전 프레임의 SNR이 3보다 크면, 평활화 대역 에너지, 이전 프레임의 추정 잡음의 1.00547배 및 미리 설정된 추정 잡음의 최대값 중 최소값을 취하며, 이전 프레임의 SNR이 3보다 작으면, 평활화 대역 에너지, 이전 프레임의 추정 잡음 및 미리 설정된 추정 잡음의 최대값 중 최소값을 취한다. 또한, 이렇게 갱신된 추정 잡음의 값이 미리 설정된 최소값보다 작으면, 추정 잡음은 미리 설정된 추정 잡음의 최소값을 취한다.

- <50> 음성 통화 상황에서는, 통화자의 음성 신호에는 끊어지는 구간, 예컨대 수화자가 이야기하는 구간 혹은 통화자가 말을 하는 중에 잠깐 쉬는 구간이 반드시 포함되게 되며, 이러한 구간의 영향으로 상기 추정 잡음 갱신 알고리즘에 의한 추정 잡음은 줄어들게 된다.
- <51> 이에 반해, 오디오 신호의 경우, 이러한 신호가 끊어지는 구간이 없는 경우가 많으므로 추정 잡음 갱신 알고리즘에 의한 추정 잡음은 시간의 경과에 따라 처음에 설정된 값보다 1.03배 혹은 1.00547배 등으로 점점 커지게 되며, 다만 추정 잡음의 크기가 평활화 대역 에너지보다 커지게 되는 경우에만 추정 잡음의 다음 갱신값이 줄어들게 된다. 이에 따라, 평활화 대역 에너지가 비교적 일정한 범위로 유지된다면, 시간의 경과에 따라 추정 잡음이 점점 커지고 그 결과 전송률 결정 임계치의 값이 점점 증가하게 되며(수학식 1 참조), 따라서 같은 대역 에너지를 갖는 프레임이 전송률 1/8로 부호화될 확률이 점점 커지게 된다. 즉, 음악을 재생하는 시간이 길어질 경우에는 끊어짐 현상의 발생 빈도가 증가할 수 있다.
- <52> 도 2a와 2b는 각각 1분 길이의 통화 음성과 음악을 EVRC로 부호화할 경우 대역 1의 추정 잡음(도면에서는 '배경 잡음'으로 표시됨)이 시간에 따라 변화하는 추이를 보여주는 도면이다. 음성 통화의 경우에는 추정 잡음이 감소하는 구간이 여러 곳에서 발견되지만, 음악의 경우에는 추정 잡음이 계속적으로 증가하고 있는 것을 확인할 수 있다.
- <53> 도 3a와 3b는 각각 위에서 사용한 통화 음성과 음악을 EVRC로 부호화할 경우 대역 1의 두 번째 전송률 결정 임계치의 시간에 따른 변화 추이를 보여주는 도면이다. 도 2a 및 2b의 추정 잡음의 변화 양상과 유사함을 확인할 수 있다.
- <54> 수학식 3에서의 장기 예측 이득 β 는 다음 식과 같이 오차(residual)의 자기상관함수에 의해 정의된다.

<55>

【수학식 4】

$$\beta = \max \left\{ 0, \min \left\{ 1, \frac{R_{\max}}{R_{\varepsilon}[0]} \right\} \right\}$$

<56> 여기서, ε 는 예측 오차 신호(prediction residual)이고, R_{\max} 는 예측 오차 신호의 자기상관함수의 계수의 최대값이며, $R_{\varepsilon}(0)$ 는 예측 오차 신호의 자기상관함수의 0번째 계수이다.

<57> 위의 수학식 4에 따르면, 하나의 주요한(dominant) 피치가 존재하는 단선율이나 음성 신호의 경우는 β 가 큰 값을 갖고, 여러 개의 피치가 혼재하는 음악의 경우에는 β 가 작은 값을 갖는다.

<58> 수학식 4의 예측 오차 신호 ε 는 다음 식에 의해 정의된다.

<59>

【수학식 5】

$$\varepsilon[n] = s'[n] - \sum_{i=1}^{10} a_i s'[n-i]$$

<60> 여기서, $s'[n]$ 은 잡음 억제 전처리된 음성 신호이며, a_i 는 현재 프레임의 LPC 계수이다.

<61> 즉, 예측 오차 신호는 LPC계수에 의해 재구성된 신호와 원본 신호와의 오차이다.

<62> 도 4a와 도 4b는 각각 주 피치(dominant pitch) 성분이 존재하는 신호와 그렇지 않은 신호의 프레임 예측 오차 신호의 예를 나타낸 도면이다. 프레임 오차 신호(frame residual signal)는 프레임 안에 주 피치가 존재하는 경우(도 4a)에는 비교적 규칙적인 양상을, 주 피치가 존재하지 않는 경우(도 4b)에는 불규칙적인 양상을 보인다.

<63> 도 5a와 도 5b는 각각 도 4a와 도 4b의 오차 신호의 자기상관을 구한 것을 그린 도면으로서, 도면에서의 자기상관은 오차 에너지 $R_{\varepsilon}[0]$ 에 대해 정규화된 것이다. 두 그림에서 원으로 표시된 부분이 R_{\max} 의 위치이다. 수학식 4에서 살펴본 바와 같이, 오차의 정규화된 최대첨두치 자기상관계수 값은 장기 예측 이득 β 를 의미하는데, 주 피치가 존재하는 경우(도 5a)

에는 β 가 0.6792로 상대적으로 큰 값을 갖게 되고, 주 피치가 존재하지 않는 경우(도 5b)에는 β 가 0.2616으로 상대적으로 작은 값을 갖게 된다.

<64> 이제 전송률 결정에 대해서 알아본다. 두 주파수 대역에서 각기 두 임계치와 대역 에너지를 비교하여 대역 에너지가 두 임계치 모두보다 크면 전송률 1로, 두 임계치 사이라면 전송률 1/2로, 두 임계치 모두보다 작으면 전송률 1/8로 결정한다. 이렇게 두 주파수 대역에서 전송률을 결정하고, 각 주파수 대역에 대해 결정된 두 전송률들 중 큰 것을 현재 프레임의 전송률로 선택한다.

<65> 이상에서 살펴본 바를 기초로 하여, 본 발명은 음폭 압축(dynamic range compression) 및 피치 강화(pitch enhancement) 전처리를 제안한다.

<66> 도 6은 본 발명에서 제안하는 음악 전처리 과정의 개념도이다. 서버단 컴퓨터(610)에서는 입력 오디오 신호에 대해 자동 이득 제어(AGC; Automatic Gain Control)에 의한 음폭 압축(610a)과 심리 음향 모델에 근거한 피치 강화 전처리 과정(610b)이 수행되고, 기지국(Base station; 620)에서는 이렇게 전처리된 오디오 신호가 EVRC 부호기(620a)에 의해 부호화되며, 최종적으로 이동통신 단말기(630)를 통하여 복호화(630a)되어 사용자가 출력 오디오 신호를 듣게 된다.

<67> 본 발명은 도 6에서 도시된 과정 중, 서버단 컴퓨터(610)에서의 자동 이득 제어에 의한 음폭 압축 전처리(610a)와 심리 음향 모델에 근거한 피치 강화 전처리(610b)를 제안하고 있는 바, 먼저 음폭 압축 전처리에 대해 구체적으로 살펴본다.

<68> 음폭 압축 전처리는 입력 오디오 신호의 음폭이 이동 통신 시스템이 감당할 수 있는 음폭보다 지나치게 넓어서 소리가 작은 부분이 잘 들리지 않거나 큰 부분이 포화되는 경우를 해

결하는 방법이며, 이를 통해 이동통신 단말기의 스피커 특성에 맞는 최적 음질의 오디오를 제공할 수 있게 된다. 음성 통화 상황이라면 에너지가 작은 프레임은 잡음 프레임으로 간주할 수 있지만, 음악의 경우에는 에너지가 작다고 해서 잡음 프레임으로 생각할 수는 없다. 또, 이동 통신 시스템이 감당할 수 있는 음폭은 매우 좁고, 또한 EVRC의 전송률 결정 알고리즘은 에너지가 작은 프레임을 잡음으로 인식하는 경향이 있으므로, 넓은 음폭을 가진 음악의 경우 음악의 에너지가 아주 작은 부분에서 끊김 현상이 초래될 수 있다. 따라서, 음폭이 넓은, 예를 들어 CD 음질 오디오에 대해서는 음폭 압축을 수행할 필요가 있다. 본 발명에서는 음폭 압축을 AGC에 의해 구현한다.

<69> AGC는 시간적으로 어택 시간(attack interval) 이후까지 신호를 미리 보고 현재 신호의 이득을 조절하는 방법이다. 예를 들어, 사람이 주변 사람들을 고려해야 하는 아파트에서 음악 소리가 너무 크다면 주변 사람에게 피해를 줄 수 있기 때문에 큰 소리가 재생된 후에 볼륨을 낮추게 될 것이다. AGC는 이러한 과정을 자동으로 수행하게 해 주며, 또한 사전에 볼륨 조절을 해 줄 수 있는 방법이다. 또 다른 예로서, 스피커마다 음폭(dynamic range)이 다른데, AGC 처리를 하지 않고 음악을 재생할 때 재생하는 스피커가 처리할 수 없는 음폭을 재생해야 된다면 그 스피커에서 나오는 소리는 포화(saturation)되게 될 것이다. 따라서, 스피커, 이어폰, 혹은 휴대 전화 등의 재생 장치의 특성에 맞는 AGC 처리는 필수적이다.

<70> 휴대폰에 대해 가장 적절한 음질을 보장하기 위해서는 휴대폰의 음폭을 조사하여 AGC를 수행해 주면 이상적이지만, 휴대폰 제조사마다 스피커 특성이 달라 모든 휴대폰에 대해 최적화된 AGC의 설계는 불가능하다. 따라서, 모든 휴대폰에 대해 범용으로 적용될 수 있는 적절한 수준의 AGC가 필요하다.

- <71> 도 7은 AGC를 수행하기 위한 블록도이다. 먼저, 신호 $s[n]$ 을 이용하여 정방향 신호 레벨 $l_f[n]$ 과 역방향 신호 레벨 $l_b[n]$ 을 계산하여, 최종 신호 레벨 $l[n]$ 을 계산한다. $l[n]$ 이 계산되면 이를 이용하여 신호별 처리 이득 $G[n]$ 을 계산하고, 이렇게 계산된 $G[n]$ 에 신호 $s[n]$ 을 곱하여 출력 신호 레벨 $y[n]$ 을 계산해 낸다.
- <72> 이하에서 도 7의 각 블록이 수행하는 기능을 도면을 사용하여 설명한다.
- <73> 도 8은 샘플 오디오 신호 $s[n]$ 에 대해 구한 신호 레벨 $l[n]$ 을 도시한 것이다. $l[n]$ 을 결정할 때, 정방향의 지수형 감쇄(ATTACK)와 역방향의 지수형 감쇄(RELEASE)를 이용하는데, 두 감쇄를 이용하여 신호를 어떻게 처리하느냐에 따라서 신호 레벨의 모양이 달라진다. 도 8에서 L_{max} , L_{min} 은 AGC 처리후 신호가 가질 수 있는 최대 및 최소값을 나타낸다.
- <74> 시각 n 에서의 신호 레벨은 릴리즈(RELEASE)의 구현을 위한 정방향 신호 레벨의 계산과 어택(ATTACK)의 구현을 위한 역방향 신호 레벨의 계산에 의해 이루어진다. 따라서, 상기 지수형 감쇄의 특성을 결정짓는 "지수함수"의 시간 상수는 정방향의 경우 릴리즈 시간(RELEASE TIME)으로, 역방향의 경우 어택 시간(ATTACK TIME)으로 정의된다. 이하에서 도 9 및 도 10을 이용하여 각각 정방향 신호 레벨 계산과 역방향 신호 레벨 계산 방법을 설명한다.
- <75> 먼저, 도 9를 사용하여 정방향 신호 레벨 계산 방법을 설명한다.
- <76> 제1 단계로 현재 첨두치(current peak)와 현재 첨두치 인덱스(current peak index)를 각각 0으로 초기화하고, 정방향 신호 레벨 $\{l_f[n]\}$ 을 신호 값의 절대값 $\{|s[n]|\}$ 으로 초기화한다.
- <77> 제2 단계로 현재 첨두치와 현재 첨두치 인덱스를 갱신한다. $|s[n]|$ 이 현재 첨두치 $p[n]$ 보다 크다면 $p[n]$ 을 $|s[n]|$ 으로, 현재 첨두치 인덱스 $i_p[n]$ 을 n 으로 갱신한다.

```

<78>      if ( |s[n]| > p[n] ) {
                p[n] = |s[n]|
                ip[n] = n
      }

```

<79> 제3 단계로 감쇄된 현재 첨두치(decayed current peak)를 계산한다. 감쇄된 현재 첨두치 $p_d[n]$ 은 $p[n]$ 값을 시간의 경과에 따라 지수적으로 감쇄시킨 값이다.

```

<80>      pd[n] = p[n] x exp(-TD / RT)

```

【수학식 6】 $TD = n - i_p[n]$

<81> 여기서, RT는 릴리즈 시간(Release Time)이다.

<82> 제4 단계로 정방향 신호 레벨을 결정한다. $p_d[n]$ 과 $|s[n]|$ 의 두 값 중 큰 값으로 정해진다.

```

<83>      【수학식 7】  $l_f[n] = \max(p_d[n], |s[n]|)$ 

```

<84> 제5 단계로 제2 단계에서 제4 단계의 과정을 $n = 0$ 에서 시작하여 시간을 증가시키며 반복하여 정방향 신호 레벨 $\{l_f[n]\}$ 을 얻는다.

<85> 다음으로, 도 10을 참조하면서, 역방향 신호 레벨 계산 방법을 설명한다.

<86> 제1 단계로 현재 첨두치를 0으로, 현재 첨두치 인덱스를 어택 시간 AT로 초기화하고, 역방향 신호 레벨 $\{l_b[n]\}$ 을 신호 값의 절대값 $\{|s[n]|\}$ 으로 초기화한다.

<87> 제2 단계로 현재 첨두치와 현재 첨두치 인덱스를 갱신한다. n 부터 $n + AT$ 까지의 시간 윈도우(time window) 안에서 신호의 최대값을 찾아서, 그 값을 현재 첨두치 $p[n]$ 으로, 그 시간 인덱스를 $i_p[n]$ 으로 갱신한다.

<88> $p[n] = \max(\{|s[]|\})$

【수학식 8】 $i_p[n] = (\text{최대값을 갖는 } \{|s[]|\} \text{의 인덱스})$

<89> 여기서, s 의 인덱스는 n 부터 $n + AT$ 까지이다.

<90> 제3 단계로 감쇄된 현재 침투치를 계산한다.

<91> $p_d[n] = p[n] \times \exp(-TD / AT)$

【수학식 9】 $TD = i_p[n] - n$

<92> 제4 단계로 역방향 신호 레벨을 결정한다. 감쇄된 현재 침투치 $p_d[n]$ 과 $|s[n]|$ 의 두 값 중 큰 값으로 정해진다.

<93> 【수학식 10】 $l_b[n] = \max(p_d[n], |s[n]|)$

<94> 제5 단계로 제2 단계에서 제4 단계의 과정을 $n = 0$ 에서 시작하여 시간을 증가시키며 반복하여 역방향 신호 레벨 $\{l_b[n]\}$ 을 얻는다.

<95> 최종 신호 레벨 $\{l[n]\}$ 은 각 시간 인덱스에 대해 정방향 신호 레벨과 역방향 신호 레벨의 최대값으로 정의된다.

<96> 【수학식 11】 $l[n] = \max(l_f[n], l_b[n])$ for $t = 0, \dots, t_{\max}$

<97> 여기서, t_{\max} 는 시간의 최대 인덱스이다.

<98> 신호 레벨을 계산할 때, 어택(ATTACK)과 릴리즈(RELEASE) 시간 값을 적절하게 설정해야 매체의 특성에 최적화된 음질을 얻을 수 있다. 어택과 릴리즈 시간의 합이 너무 짧다면(즉, 20ms미만이라면), 주파수가 '1000 / (어택시간 + 릴리즈 시간)'의 진동("vibrating") 왜곡이 들릴 수 있다. 즉, 어택을 5ms, 릴리즈를 5ms로 설정한다면 100Hz의 진동 왜곡이 들리게 된다

. 따라서, 어택과 릴리즈 시간의 합이 최소 30ms는 넘어야 AGC로 인한 진동 왜곡을 방지할 수 있다.

<99> 예를 들어, 어택을 느리게 하고 릴리즈를 빠르게 하면 더 넓은 음폭을 얻을 수 있다.

릴리즈시간이 짧으면 출력 음악 신호는 고주파 성분이 억제되어 보다 밋밋하게 된다. 하지만, 릴리즈 시간을 굉장히 빠르게 한다면(음악의 특성에 따라 굉장히 빠르다는 정도는 다름) AGC 처리 출력 음악은 입력 파형의 저주파 성분을 따라가게 되어서, "기본(fundamental)" 성분이 억제되고 "기본" 성분을 특정한 종류의 "화음 왜곡(harmonic distortion)"으로 대체하게 될 수도 있다. 여기서, "기본" 성분이란 사람의 귀가 인식하는 가장 중요한 주파수 성분으로서 "피치(pitch)"와 같은 의미를 갖는다.

<100> 드럼과 같은 타악기의 경우, 타악기 소리가 충분히 강조되게 하려면 어택시간을 길게 하는 것이 좋으며, 사람의 음성이 들어간 부분에 대해서는 어택 시간을 짧게 처리해주어 불필요하게 음성의 시작 앞부분의 이득이 줄어드는 것을 방지하는 것이 좋다. 어택과 릴리즈 시간을 결정하는 것은 AGC 처리에 있어서 음질을 보장하기 위해 중요한 사항이며, 또한 음악의 성질에 따라 결정된다.

<101> 다음으로 심리 음향 모델에 근거한 피치 강화 전처리에 대해 살펴본다.

<102> 피치 강화 전처리의 핵심은 수학적 4의 장기 예측 이득 β 를 증가시켜서 변형된 신호가 EVRC 부호화 과정에서 전송률 1로 부호화될 수 있도록 하는 것이다. 다만, 이 때의 전처리 과정은 심리 음향 모델에 근거함으로써 체감 음질의 왜곡을 최소화해야 한다.

<103> 먼저 본 발명에서 사용한 심리 음향 모델을 살펴본 후, 구체적인 피치 강화 전처리 과정을 살펴보기로 한다.

<104> 심리 음향 모델이란 인간의 귀가 가지는 지각 특성을 모델링한 것으로서 오디오의 압축 전송에 널리 이용되어 왔으며, 구체적으로는 사람이 오디오를 청취할 경우 오디오 신호의 전부를 인식하는 것이 아니라, 차폐 효과(masking effect)에 따라 신호의 일부만을 인식하게 되는 것을 이용한 것이다. 본 발명에서는 여러 심리 음향 모델 중 오디오의 압축에 널리 사용되는 최대 가청 임계치(ATH: Absolute Threshold of Hearing), 임계 대역(Critical Band), 동시 차폐와 차폐의 확산 현상 (Simultaneous masking and the spread of masking) 모델들을 사용하였는데, 이 모델들은 MP3(MPEG I Audio layer 3)에서도 사용된다. 본 발명에서 사용한 심리 음향 모델들을 구체적으로 살펴보면 다음과 같다.

<105> 최대 가청 임계치는 잡음이 없는 환경에서 청자가 단일 주파수 음(pure tone)을 감지하는데 필요한 최소 에너지이다. 이는 Fletcher의 실험에서 처음으로 확인되었으며, Terhardt가 다음과 같이 비선형 식으로 정량화하였다.

<106> 【수학식 12】
$$T_q(f) = 3.64(f/1000)^{-0.8} - 6.5e^{-0.6(f/1000-3.3)^2} + 10^{-3}(f/1000)^4 \text{ (dB SPL)}$$

<107> 여기서, SPL(Sound pressure level)은 절대 음압을 의미한다.

<108> 주파수에 따른 최대 가청 임계치 그래프는 도 8과 같다.

<109> 임계 대역폭(critical bandwidth)은 주관적 반응이 급격히 변하는 대역폭으로 정의될 수 있다. 두 차폐 톤(masking tone)들 사이에 있는 협대역 잡음 음원에 대한 감지 임계치(detection threshold)는 두 톤의 간격이 임계 대역폭 사이에 있다면 일정하게 유지되며, 두 톤의 간격이 이 대역폭을 넘으면 임계치는 급격히 감소한다(도 9(a), 9(c)). 즉, 두 차폐 톤 간의 간격이 임계 대역폭보다 작을 경우에는 감지 임계치가 일정하게 유지되어 그 임계치를 넘는 잡음만이 들리게 되나, 두 차폐 톤 간의 간격이 임계 대역폭을 넘으면 그 임계치가 점점 감

소하게 되어 더 작은 잡음도 들리게 된다. 마찬가지로, 차폐 잡음(masking noise)에 대한 톤의 실험에서도 같은 양상이 나타난다(도 9(b), 9(d)).

<110> 이와 같은 인간의 청각 특성들을 고려할 때, 평균적인 청자에 대하여 주파수별 임계 대역폭은 다음과 같이 정량화 될 수 있다.

<111> **【수학식 13】** $BW_c(f) = 25 + 75 [1 + 1.4(f/1000)^2]^{0.69} \text{ (Hz)}$

<112> $BW_c(f)$ 는 연속함수이지만, 실제 응용에 있어서는 인간의 귀가 위 식을 만족시키는 대역통과 필터의 불연속적 집합으로 이루어져 있다고 생각하는 것이 편리하다. 두 임계 대역 사이의 간격은 보통 1 bark로 칭한다. 다음의 수학식 14는 주파수를 bark로 변환하는 식이다.

<113> **【수학식 14】** $z(f) = 13 \arctan(0.00076f) + 3.5 \arctan [(f/7500)^2] \text{ (bark)}$

<114> 차폐(masking)는 어떤 음원의 존재로 인해 다른 음원이 들리지 않게 되는 현상을 말한다. 동시 차폐는 하나의 임계 대역 내에서 관찰되는 주파수 영역의 차폐 현상을 말한다. 동시 차폐는 두 가지 종류가 있는데, 하나는 톤-잡음-차폐(tone-masking-noise)이며, 또 다른 하나는 잡음-톤-차폐(noise-masking-tone)이다. 톤-잡음-차폐는 임계 대역의 중심에 위치한 톤이 임계 대역폭보다 작은 대역폭을 갖는 잡음을 차폐하는 현상을 말하며, 이 때 잡음의 스펙트럼은 차폐 톤의 세기와 관련된 예측가능한 임계치 곡선(predictable threshold curve)보다 아래에 있어야 한다. 잡음-톤-차폐는 톤-잡음-차폐에서 차폐자(masker)와 피차폐자(maskee)의 역할을 서로 바꾼 것이다. 즉, 강한 잡음 차폐자나 톤 차폐자는 임계 대역에 위치한 기저막(basilar membrane: 인간의 귀의 일 기관으로서, 기저막을 따라 주파수-위치 변환이 일어난다는 것이 알려져 있다)에 충분한 강도의 자극을 생성하여, 약한 신호가 감지되지 못하도록 한다.

- <115> 대역간 차폐(inter-band-masking) 역시 관찰되고 있다. 즉, 하나의 임계 대역 내에 있는 차폐자는 인접한 다른 대역의 감지 임계치에도 영향을 미치며, 이러한 효과는 "차폐의 확산 현상"으로 알려져 있다.
- <116> 이제 구체적으로 본 발명에서 제안하는 피치 강화 전처리에 대해서 살펴보기로 한다.
- <117> 도 10은 한 프레임의 오디오 신호의 피치를 강화하기 위한 블록도이다. 입력 오디오 신호를 주파수 영역으로 옮긴 후(1010, 1020), 심리 음향 모델에서 얻은 전역 차폐 임계(GMT: Global Masking Threshold) 곡선을 이용하여(1030, 1040) GMT 곡선 아래 부분을 억제하고 (1050) 오차 첨두치를 강화하면(1060, 1070) 장기 예측 이득 값이 커지므로, 음질의 변형을 최소화하면서 EVRC 전송률 결정 알고리즘에서 비롯되는 끊김 현상을 줄일 수 있다.
- <118> 앞서 살펴본 바와 같이, 전송률 결정 임계치는 주로 β 가 작은 값으로 오랜 시간 지속될 경우(연속 8 프레임 이상 동안 β 가 0.3 미만일 경우) 증가하게 되며, 평활화 대역 에너지가 기존의 임계치보다 작을 경우 평활화 대역 에너지에 맞추어 감소하게 된다. 여기서, β 는 오차의 최대 자기상관과 오차 에너지의 비로 결정되며, 프레임 내에 주 피치가 존재할 경우 β 는 크게 되고, 그렇지 않을 경우 β 는 작게 된다.
- <119> 이와 같은 전송률 결정 임계치의 증감 메커니즘은 이동 통신에서 음성을 주로 전송할 경우에는 적절하다. 사람의 음성 가운데 유성음 부분은 주 피치가 존재하여 β 가 커지게 되고, 그 결과 음성 정보는 높은 전송률로 부호화되어 전송되며, 반면에, 묵음 구간의 프레임들은 배경 잡음만을 포함하므로 대역 에너지가 작은 값이 되고, 그 결과 전송률 결정 임계치가 줄어들게 된다. 따라서, 음성 통화의 경우에는 전송률 결정 임계치가 배경 잡음의 정도에 따라 적절하게 유지될 수 있다.

- <120> 그러나, 음악에는 묵음 구간이 존재하지 않으므로, 전송률 결정 임계치는 점차 증가하는 양상을 보이게 된다. 음악이 단선율로서 주 피치가 항상 존재하고 대역 에너지의 시간적 변화가 일정하지 않을 경우 β 가 크기 때문에 전송률 결정 임계치가 거의 증가하지 않을 것이지만, 다양한 화음이 포함된 일반적인 음악은 이러한 성질을 만족하기 힘들다.
- <121> 따라서, 본 발명은 이러한 부작용을 완화하기 위한 방법으로 음질의 손상을 최소화하면서 β 를 증가시키는 방법을 사용한다. β 를 증가시키기 위해서는 오차의 자기상관의 최대값 (R_{\max})을 증가시키고 오차 에너지($R_{\varepsilon}[0]$)를 감소시켜야 한다. 이를 위하여 각 프레임 오디오 샘플에 대해 "다중 톤 노치 필터링(MTNF: Multi-Tone Notch Filtering; 1050)"과 "오차 침두치 강화(1070)"의 두 단계로 전처리를 수행한다.
- <122> 이 두 과정은 오차 영역이 아닌 주파수 영역에서 이루어지는데 그 이유는 다음과 같다. 첫번째 이유는 다중 톤 노치 필터링 과정은 심리 음향 모델에 기반하는데, 심리 음향 모델의 GMT 계산은 주파수 영역에서 이루어지기 때문이다. 두번째 이유는 오차 영역에서의 전처리는 신호에 보다 큰 왜곡을 줄 수 있기 때문이다. 오차는 원본 신호에 비하여 파형이 복잡하며, 특히 위상이 다를 수 있어서 오차 영역에서 전처리한 신호는 주파수 영역에서 전처리한 신호에 비하여 파형이 복잡해져 오디오의 왜곡이 크다.
- <123> 주파수 영역에서 전처리를 수행해도 되는 것은 다중 톤 노치 필터링과 오차 침두치 강화 전처리의 관점에서는 주파수 및 오차 영역에서의 전처리가 같은 의미를 갖기 때문이다. EVRC 오차는 수학식 5에 정의된 바와 같이 10개의 이전 샘플들로부터 선형 예측이 불가능한 성분들을 갖는 신호이며, 대략적으로 여기 신호(excitation signal)로 생각할 수 있다. EVRC가 계산할 수 있는 피치 간격은 최소 20 샘플이므로 10개의 계수를 갖는 선형 예측으로는 피치 성분을 결코 예측할 수 없다. 따라서, 오차는 피치 성분들을 주로 포함하며 비 피치 성분들을 적

게 포함하는 신호라고 볼 수 있다. 다중 톤 노치 필터링 과정은 주파수 영역에서 주요 주파수 성분들을 그대로 둔 채 비 주요 성분들만을 억제하므로 오차 신호에서도 비 주요 성분들만 억제하는 효과를 갖는다. 그리고, 오차 침투치 강화 과정 역시 피치 성분들을 강화하는 과정이므로 주파수 영역에서의 강화 효과가 오차 영역에서 그대로 나타나게 된다.

<124> 먼저, 다중 톤 노치 필터를 사용한 전처리(1010 - 1050)에 대해서 구체적으로 살펴보기로 한다. 전송률 결정 임계치를 작게 유지하려면 β 를 증가 시켜야 하며, 이를 위해서는 R_{\max} 를 증가시키거나 $R_{\varepsilon}[0]$ 를 감소시켜야 하는데, 다중 톤 노치 필터링은 $R_{\varepsilon}[0]$ 를 감소시키는 역할을 수행한다. 다중 톤 노치 필터를 이용한 전처리 과정에서 체감 음질의 손상을 최소화 하기 위하여 심리 음향 모델의 전역 차폐 임계치를 얻어 주파수 영역에서 GMT 곡선 아래 부분의 성분들을 선택적으로 억제한다.

<125> 본 발명에서 GMT를 얻는 과정은, 종래의 방법과 그 원리는 동일하나 전화에서 사용되는 주파수 대역인 8kHz에 적합하도록 적절히 계산 과정을 변경한 것으로서 다음과 같다.

<126> (1) 주파수 분석 및 SPL 정규화(1010 - 1030)

<127> 먼저, 입력 음악 신호(8kHz, 16bit PCM)를 EVRC의 프레임 크기인 160개씩으로 나눈 후, 고속 푸리에 변환(FFT; Fast Fourier Transform)의 수행을 위해 각 프레임의 160 샘플에 96개의 0을 덧붙여 샘플 개수가 256이 되도록 한다(1010의 제로 패딩 과정). 각 프레임에서 입력 오디오 샘플 $s[n]$ 을 FFT 길이 $N(=256)$ 과 샘플당 비트수 b 에 대해 다음식과 같이 정규화한다(1010의 정규화 과정).

<128>
$$x[n] = \frac{s[n]}{N \times 2^{b-1}}$$

【수학식 15】

<129> 정규화된 입력 신호 $x[n]$ 에 대해 FFT를 수행(1020)하여 다음 수학적 식 16과 같이 추정 전력 스펙트럼 밀도(power spectral density(PSD) estimate) $P[k]$ 를 얻는다(1030).

<130> **【수학적 식 16】** $P[k] = 90 + 20 \log_{10} |X[k]|$ (dB SPL)

<131> 여기서, $X[k]$ 는 $x[n]$ 의 이산 푸리에 변환(DFT; Discrete Fourier Transform)이다.

<132> 심리 음향 분석 동안에는 실제의 재생 기준값(playback level)에 대해 알 수 없으므로 위의 정규화 과정이 실제 SPL을 추정하는 데 사용된다.

<133> (2) 톤과 잡음 차폐자의 확인

<134> 다음 식 17을 만족하는 주파수 성분은 톤 집합(tonal set) S_T 에 추가된다.

<135> 음식에서 계산된다.

<136> **【수학적 식 17】** $S_T = \{P[k] \mid P[k] > P[k \pm 1], P[k] > P[k \pm 5] + 7\text{dB}\}$

<137> 즉, 배경 잡음보다 높은 전력 값을 갖는 주파수 성분은 톤 집합에 추가된다.

<138> 톤 집합 S_T 의 스펙트럼 첨두치들로부터 톤 차폐자(Tone masker) $P_{TM}[k]$ 는 다음식에서 계산된다.

<139> **【수학적 식 18】**
$$P_{TM}[k] = 10 \log_{10} \sum_{j=-1}^1 10^{0.1P(k+j)} \text{ (dB)}$$

<140> 톤 차폐자의 ± 5 범위 내에 있지 않은 각 임계 대역에 대해 다음식과 같이 잡음 차폐자(Noise masker) $P_{NM}[\bar{k}]$ 가 정의된다.

<141> **【수학적 식 19】**
$$P_{NM}[\bar{k}] = 10 \log_{10} \sum_j 10^{0.1P(j)} \text{ (dB)}$$

$$\forall P[j] \notin \{P_{TM}[k, k \pm 1, k \pm \Delta_k]\}$$

<142> 여기서, \bar{k} 는 다음식으로 정의되는 임계 대역 내의 스펙트럼 선(spectral line)의 기하·평균(geometric mean)이다.

<143>

$$\bar{k} = \left(\prod_{j=l}^u k_j \right)^{1/(l-u+1)}$$

【수학식 20】

<144> 여기서, l과 u는 각 임계 대역의 스펙트럼 경계(spectral boundary)의 하한과 상한값이다.

<145> (3) 차폐자의 재구성

<146> 차폐자의 개수를 다음 두 가지 원칙에 의하여 줄인다. 먼저, 최대 가청 임계치를 넘지 않는 톤 혹은 잡음 차폐자를 제외시킨다. 다음으로, 0.5 bark 너비의 윈도우를 움직여서 그 윈도우 안에 두 개 이상의 차폐자가 들어온다면 제일 큰 값을 갖는 차폐자만 남기고 나머지를 제외시킨다.

<147> (4) 개별 차폐 임계치(individual masking threshold)의 계산(1040)

<148> 각 차폐 임계치는 j번째 주파수에 위치한 톤 혹은 잡음 차폐자에 의한 i번째 주파수에서의 차폐 임계치를 의미한다. 다음식에 의하여 톤 차폐자 임계치(tonal masker threshold)가 정의된다.

<149> 【수학식 21】 $T_{TM}[i, j] = P_{TM}[j] - 0.275z[j] + SF[i, j] - 6.025 \text{ (dB SPL)}$

<150> 여기서, z[j]는 j번째 주파수의 bark를, SF[i, j]는 확산 함수(spreading function)로서 기저막 확장 함수(basilar spreading function)를 근사적으로 모델링한 식이다.

<151> 잡음 차폐자 임계치(noise masker threshold)는 다음식과 같이 정의된다.

<152> 【수학식 22】 $T_{NM}[i, j] = P_{NM}[j] - 0.175z[j] + SF[i, j] - 2.025 \text{ (dB SPL)}$

<153> (5) 전역 차폐 임계치(GMT : global masking threshold)의 계산(1040)

<154> GMT $T_{GM}[i]$ 는 다음식에 의해 정의된다.

<155>

$$T_{GM}[i] = 10 \log_{10} \left(10^{0.1r_q[i]} + \sum_{l=1}^L 10^{0.1r_{tm}[i,l]} + \sum_{m=1}^m 10^{0.1r_{mw}[i,m]} \right) \text{ (dB SPL)}$$

【수학식 23】

<156> 여기서, L과 M은 각각 톤과 잡음 차폐자의 개수이다.

<157> 이상에서 구한 GMT 곡선을 이용해 주파수 영역에서 주파수 성분이 GMT 아래에 있는 대역들을 억제함으로써, 주 피치 성분은 그대로 유지한 채 $R_{\varepsilon}[0]$ 를 감소시키는 것이 가능하다. GMT 아래에 있는 부분을 가장 극단적으로 억제하는 방법은 모두 0으로 처리하는 방법이 있겠지만, 그렇게 할 경우에는 시간 영역 앨리어싱 (time-domain aliasing)이 발생할 수 있으므로, 이러한 문제를 최소화하기 위해 코사인 함수를 사용한 억제 방법을 선택하였다. 이 방법에 사용된 주파수 영역 필터를 다중 톤 노치 필터라고 부르기로 한다. 아래는 다중 톤 노치 필터를 이용하여 음악 신호를 전처리하는 과정(1050)이다.

<158> 주파수 성분이 GMT 아래에 있는 대역들을 찾아서 $\{MB_i\}$ 라고 표시한다.

<159>

$$MB_i = (l_i, u_i)$$

<160> 여기서, l_i 와 u_i 는 i 번째 대역의 상한과 하한 주파수이다.

<161> 대역 MB_i 에 적용되는 다중 톤 노치 필터 함수 $F[k]$ 는 다음식과 같다.

<162>

$$F[k] = \begin{cases} \frac{1-\alpha}{2} \cos \frac{2\pi(k-l_i)}{u_i-l_i} + \frac{1+\alpha}{2}, & \text{for } k \in MB_i \\ 1, & \text{for } k \notin MB_i \end{cases}$$

【수학식 24】

<163> 여기서, k 는 주파수 번호를 말하며, α 는 억제 상수로 0과 1 사이의 값을 취하며 작을수록 억제 강도가 증가한다. 여러 음악에 대한 실험을 통해 억제 상수 α 의 값을 0.001로 택하였다.

<164> 위에서 구한 다중 톤 노치 필터 함수를 정규화된 입력 신호 $x[n]$ 의 이산 푸리에 변환 (DFT)인 $X[k]$ 에 곱하여 처리된 출력인 $\tilde{X}[k]$ 를 다음식과 같이 얻는다.

<165> **【수학식 25】** $\tilde{X}[k] = X[k] \times F[k] \text{ for } 0 \leq k < 256$

<166> 이상의 과정에 의해 GMT를 넘는 주파수 성분들은 강화되고, GMT를 넘지 않는 주파수 성분들은 억제되는 효과가 발생되며, 그 결과 오차 에너지 값이 감소된다. 도 11은 입력 신호에 대해 다중 톤 노치 필터 함수를 적용시킨 결과이다. 다중 톤 노치 필터에 의해서 처리된 스펙트럼이 원본 스펙트럼에 비하여 주요 주파수 성분이 강화되고 차폐 임계 곡선 아래 부분이 억제된 것을 확인할 수 있다.

<167> 도 12는 다중 톤 노치 필터를 샘플 오디오에 적용시켰을 경우, 오차 에너지 $R_e[0]$ 의 변화를 나타낸 도면이다. 도면에서 굵은 실선으로 표시된 것이 다중 톤 노치 필터링 처리 후의 오차 에너지인데, 전 프레임에 대해 점선으로 표시된 처리 전의 오차 에너지보다 작아진 것을 확인할 수 있다.

<168> 다음으로, 오차 침두치 강화 전처리 방법(1060, 1070)에 대하여 구체적으로 살펴보기로 한다. 다중 톤 노치 필터 함수로 처리된 프레임 신호 샘플들을 EVRC 부호기에 입력시켜 피치 D 를 구한다(1060). 여기서, D 는 시간 영역 오차 자기상관에서 침두치들이 나타나는 샘플의 간격을 의미한다. 자기상관과 전력 스펙트럼 밀도(Power Spectral Density)는 서로 푸리에 변환 쌍(Fourier transform pair) 관계에 있으므로, 오차 자기상관이 시간 영역에서 D 간격으로 침두치가 나타나는 신호라면, 오차의 스펙트럼은 N/D 간격으로 침두치가 나타나게 된다. 따라서, 주파수 영역에서 N/D 간격의 신호들을 강화하면 시간 영역에서 D 간격의 성분들이 강화되고, 이에 따라 D 간격의 오차 성분들이 커져서 β 값이 증가하는 효과를 얻을 수 있다.

<169> 오차 침투치 강화의 성능은 크게 두 가지 요소, 즉, 다중 톤 노치 필터 적용 여부 및 그 강화 방법에 의존한다.

<170> (1) 다중 톤 노치 필터 적용 여부

<171> 주파수 영역에서 주 피치가 거의 나타나지 않는 헤비 메탈 음악의 경우, 다중 톤 노치 필터를 적용시키면 주파수 영역 성분들의 크기가 불규칙적으로 억제되므로 N/D 간격으로 강화 하더라도 β 의 증가 효과가 크지 않다. 따라서, 헤비 메탈 음악의 경우에는 다중 톤 노치 필터를 적용시키지 않는 편이 더 좋다.

<172> (2) 선택된 주파수 성분의 강화 방법(1070)

<173> 본 발명에서 선택된 주파수 성분을 강화하는 데에는 두 가지 방법이 사용된다. 첫 번째 방법은 각 성분에 대해 GMT까지만 강화하는 방법이고, 두 번째 방법은 일정 배수만큼 곱하여 강화하는 방법이다.

<174> 첫 번째 방법은 다음식과 같이 강화해야 할 주파수 성분들 중 심리 음향 모델에서 얻은 GMT 곡선 아래에 있는 성분들만 GMT까지 강화하는 것이다.

<175>

$$Y[k] = \begin{cases} T_{GM}[k], & \text{for } k = \lfloor l \times N/D \rfloor \text{ and } \tilde{X}[k] < T_{GM}[k] \\ \tilde{X}[k], & \text{otherwise} \end{cases}$$

【수학식 26】

<176> 이와 같은 방법을 적용할 경우 음질의 차이는 거의 없지만, β 는 그다지 많이 증가되지 않는다. 따라서, 이러한 방법을 통해 끊김 현상을 해결할 수 있는 노래는 제한적이다.

<177> 두 번째 방법은 프레임 주파수 영역 성분들에 대하여 피치 강화 응답 $H[k]$ 를 곱해주는 방법이다.

<178>

$$Y[k] = \tilde{X}[k] \times H[k]$$

【수학식 27】

$$H[k] = \begin{cases} 1, & 0 \leq k < \lfloor N/p \rfloor \\ \eta \cos\left(\frac{2\pi k}{N/p}\right) + (1-\eta), & \lfloor N/p \rfloor \leq k < N \end{cases}$$

<179>

위의 수학식 27에서 p 는 프레임별로 계산된 피치이며, k 는 이산 푸리에 변환 주파수 번호로 0에서 255사이의 값을 가지며, $Y[k]$ 와 $\tilde{X}[k]$ 는 각각 출력 주파수 응답 및 다중 톤 노치 필터를 거친 정규화된 프레임 오디오 신호 $x[n]$ 의 주파수 응답을 나타낸다.

<180>

피치 강화 응답 $H[k]$ 계산식을 살펴보면, $H[k]$ 는 주파수 영역에서 나타나는 기본 피치 주파수의 배수 주파수들의 성분들은 1의 값을 갖게 되고, 각 배수 주파수들 사이 주파수들의 성분들은 1보다 작은 값들을 갖게 된다. 즉, 피치 하모닉스 성분들은 기존 값을 유지하며, 피치 하모닉스가 아닌 주파수의 성분들은 상대적으로 억제되게 된다. 피치 하모닉스 성분들이 상대적으로 뚜렷해짐에 따라서 시간 영역에서 피치 성분들이 뚜렷해지고, 그에 따라 전체적으로는 장기 예측 이득이 증가하게 된다.

<181>

끝으로, m 번째 프레임의 정규화된 주파수 영역 신호 $X_m[k]$ 에 대해 피치 강화 과정을 적용시킨 주파수 영역 신호를 $Y_m[k]$, $Y_m[k]$ 를 역-고속 푸리에 변환 (IFFT; Inverse Fast Fourier Transform)하여 역정규화한 신호를 $y'_m[n]$ 이라고 하고, 출력값을 계산해 본다(1080).

<182>

피치 강화 과정은 주파수 영역 필터를 적용시키는 과정이므로 선형 컨볼루션 (linear convolution)을 하기 위해 오버랩 & 애드(overlap and add) 과정이 필요하다. 본 발명에서 오버랩 & 애드는 다음 식과 같이 구현된다.

<183>

【수학식 28】

$$y_m[n] = \begin{cases} y'_m[n] + y'_{m-1}[n+L], & \text{for } 0 \leq n < M \\ y'_m[n], & \text{for } M \leq n < L \end{cases}$$

- <184> 위의 수학적 식 28에서 $y_m[n]$ 은 m 번째 프레임의 n 번째 샘플의 출력값을 의미하고, L 은 프레임 샘플의 개수로 160이며, M 은 오버랩 버퍼의 크기로 $N - L = 96$ 이다.
- <185> 본 발명에 따른 방법으로 구현함으로써, 음악 신호의 전송률을 높일 수 있었으며, 그에 따라 EVRC에 의해 음악의 끊김 현상이 발생하는 것을 현저하게 줄일 수 있었다.
- <186> 구체적인 실험 결과는 다음과 같다. 8kHz, 16bit 모노로 샘플링된 음악 신호를 사용하여 실험하였으며, 입력 오디오 신호의 음질을 실제 전화 상황과 같게 하기 위하여 ITU-T 권고안에 정의된 대로 앨리어싱 방지 필터(anti-aliasing filter)의 주파수 응답을 200 Hz부터 3400 Hz까지 평평하게 유지하여 2 dB 내외의 편차를 갖도록 하였다.
- <187> 선정된 곡들에 대해 제안된 전처리 방법, 즉 피치 강화 전처리와 음^ㅍ폭 압축 전처리를 순차적으로 적용하였다.
- <188> 도 13의 (a) 및 (b)는 "X-Japan"의 "Silent Jealousy"라는 음악에 대하여 제안된 전처리를 수행하였을 때, 전처리에 의한 전송률 결정 임계치와 대역 에너지의 변화를 보여준다. 원본 오디오 신호(도 13의 (a))는 처음 15초 이후에 전송률 결정 임계치가 대역 에너지보다 높은 값을 유지하여 계속 끊김 현상이 나타나게 되지만, 전처리된 오디오 신호(도 13의 (b))는 전송률 결정 임계치가 대역 에너지보다 낮은 값을 유지하여 끊김 현상이 거의 나타나지 않게 된다는 것을 확인할 수 있다.
- <189> 도 14의 (a) ~ (d)는 각각 원본 오디오 신호 및 원본 오디오 신호의 EVRC 부/복호된 신호와 전처리된 오디오 신호 및 전처리된 오디오 신호의 EVRC 부/복호된 신호의 파형이다. 원본 오디오 신호(도 14의 (a))의 EVRC 부/복호한 신호(도 14의 (c))를 보면, 시간 인덱스가 100000을 지나면서부터 끊김 현상이 발생하여 끊김 현상이 계속 지속되는 것을 확인할 수

있다. 그러나, 전처리된 오디오 신호(도 14의 (b))의 EVRC 부/복호한 신호(도 14의 (d))를 보면 끊김 현상의 발생이 거의 없다는 것을 알 수 있다.

<190> 【표 2】

	원본 신호	전처리를 거친 신호
전송률 1/8의 개수	1567	29

<191> 위의 표 2는 도 14의 (a) 및 (b) 신호에 대해 각각 EVRC 부호화를 수행하였을 때, 전송률 1/8로 부호화된 프레임의 개수를 나타낸 것이다. 전처리를 거친 신호의 경우 전송률 1/8로 부호화된 프레임 개수가 현저히 줄어들었음을 확인할 수 있다.

<192> 제안된 전처리 알고리즘으로 전처리한 음악과 원본 음악에 대해 11명의 건강한 20대, 30대 피험자들을 대상으로 MOS 테스트를 실시해 보았다. MOS 테스트는 많은 수의 피험자들을 대상으로 음성 코덱에서 부/복호된 음성 신호에 대해 주관적 음성 명료도(clarity)를 평가하게 하는 방법으로 ITU-T 권고안 P.800에서 제안되었으며, 음성 코덱의 성능을 평가하는 정량적인 방법으로 널리 사용되어 왔다. 삼성 애니콜(Samsung Anycall) 단말기로 테스트하게 하였으며, 랜덤한 순서로 원본 음악과 전처리된 음악을 제공하여 각 음악에 대해 5점 척도로 절대 평가하게 하였다. 다만, 이동 통신 단말기의 음질을 평가할 때 CD음질을 그 판단 기준으로 한다면 곤란할 것이기에, 피험자들로 하여금 유선 전화상에서의 음질을 최적 음질이라고 생각하여 기준을 잡게 하였다. 5점 척도의 평가 기준은 다음과 같다.

<193> (1) bad (2) poor (3) fair (4) good (5) excellent

<194> 아래의 표 3은 MOS 테스트의 통계 분석 결과를 나타낸 것이다. 통계 분석을 위해 등분산 가정 T-test 단측 검정을 수행하였다. 통계 분석 결과를 보면, 원곡 평균이 1점대인데 비하여 본 발명에서 제안된 전처리를 한 처리곡 평균은 2 ~ 3점대로 높아져서, 처리곡이 원곡에

비해 음질이 향상되었음을 확인할 수 있다. 다만, 일부 곡의 경우 음질의 개선 효과가 그리 크지 않은 경우도 발견되는데, 그것은 음폭 압축 전처리 파라미터의 설정 과정에서 곡마다 모두 동일한 파라미터를 적용함으로써 곡의 특성에 맞게 처리되지 못한 것에 그 원인이 있을 것으로 생각된다.

<195> 【표 3】

노래제목	노래장르	원곡 평균	처리곡 평균
Slam (노바소닉)	하드 록 음악	1.727	3.000
silent jealousy (X-Japan)	하드코어 음악	1.000	2.545
Kurenai (X-Japan)	하드코어 음악	1.200	2.091
나쁜여자 (노바소닉)	하드 록 음악	1.455	2.455
스타크래프트 비창 (미상)	록 음악	1.545	3.000
소녀의 기도 (바다르체프스카)	피아노 솔로	3.000	3.273
How do I live (Le'Ann Rimes)	발라드 음악	2.727	2.727
비창 (베토벤)	피아노 솔로	1.727	2.455
Caruso (파바로티)	가곡	2.545	2.273
운명 (베토벤)	교향곡	2.091	2.727

【발명의 효과】

<196> 본 발명에 따른 전처리 방법으로 구현함으로써, 음악 신호의 전송률을 높일 수 있었으며, 그에 따라 EVRC에 의해 음악의 끊김 현상이 발생하는 것을 현저하게 줄일 수 있었다. 나아가, 휴대전화에서 듣는 음악의 음질을 향상시키는 효과도 얻을 수 있었다.

<197> 한편, 본 발명은 도면에 도시된 실시예들을 참고로 설명되었으나 이는 예시적인 것에 불과하며, 당해 기술분야에 통상의 지식을 지닌 자라면 이로부터 다양한 변형 및 균등한 타실시

예가 가능하다는 점을 이해할 것이다. 따라서, 본 발명의 진정한 기술적 보호범위는 첨부된 특허청구범위에 의해서만 정해져야 할 것이다.

【특허청구범위】**【청구항 1】**

이동 전화상의 음질을 향상시키기 위한 디지털 오디오 신호의 전처리 방법으로서,
오디오 데이터를 음폭 압축(Dynamic Range Compression) 전처리하는 단계, 및
오디오 데이터를 피치 강화(Pitch Enhancement) 전처리하는 단계를 포함하는 오디오 파일의 전처리 방법.

【청구항 2】

제1항에 있어서,
상기 음폭 압축 전처리가 자동 이득 제어(AGC)에 의한 것임을 특징으로 하는 오디오 파일의 전처리 방법.

【청구항 3】

제2항에 있어서,
상기 자동 이득 제어 처리는,
정방향으로 신호 레벨을 계산하는 단계와,
역방향으로 신호 레벨을 계산하는 단계와,
상기 계산된 정방향 신호 레벨과 역방향 신호 레벨에 기초하여 최종 신호 레벨을 생성하는 단계에 의해 신호 레벨을 계산하여 처리 신호를 생성하는 것을 특징으로 하는 오디오 파일의 전처리 방법.

【청구항 4】

제1항에 있어서,

상기 피치 강화 전처리 단계는,
다중 톤 노치 필터(MTNF: Multi-Tone Notch Filter)를 적용하는 단계와,
오차 첨두치 강화(Residual Peaks Enhancement) 단계
를 포함하는 것을 특징으로 하는 오디오 파일의 전처리 방법.

【청구항 5】

제4항에 있어서,
상기 다중 톤 노치 필터를 적용하는 단계는,
심리 음향 모델의 전역 차폐 임계치(GMT: Global Masking Threshold)를 구하여 주차수
영역에서 GMT 곡선 아래 부분의 성분들을 선택적으로 억제함으로써 수행되는 것을 특징으로 하
는 오디오 파일의 전처리 방법.

【청구항 6】

제5항에 있어서,
상기 심리 음향 모델의 전역 차폐 임계치를 구하는 과정은,
주파수를 분석하고, 절대 음압(SPL: Sound Pressure Level)을 정규화하는 단계;
톤 차폐자 및 잡음 차폐자를 계산하는 단계;
차폐자를 재구성하는 단계;
개별 차폐 임계치를 계산하는 단계; 및
전역 차폐 임계치를 계산하는 단계
를 포함하는 것을 특징으로 하는 오디오 파일의 전처리 방법.

【청구항 7】

제4항 내지 제6항에 있어서,

음악의 종류에 따라서 다중 톤 노치 필터의 적용을 선택적으로 하는 것을 특징으로 하는 오디오 파일의 전처리 방법.

【청구항 8】

제4항에 있어서,

상기 오차 침투치 강화 단계는,

심리 음향 모델에서 얻은 GMT 곡선 아래에 있는 성분들만을 GMT까지 강화하는 방법; 또는

선택된 주파수 성분들에 대해 소정의 침투치 강화 배율(PEF: Peaks Enhancement Factor)을 곱해주는 방법을 사용하는 것을 특징으로 하는 오디오 파일의 전처리 방법.

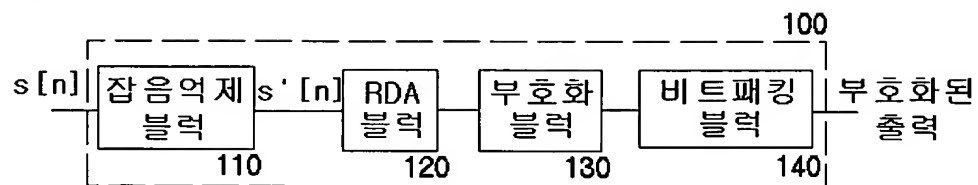
【청구항 9】

제1항 내지 제6항, 또는 제8항에 있어서,

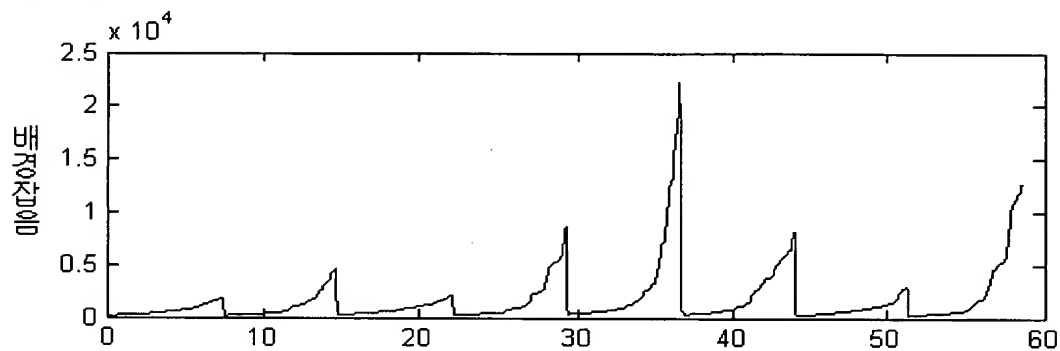
상기 전처리 과정은 심리 음향 모델에 근거하는 것을 특징으로 하는 오디오 파일의 전처리 방법.

【도면】

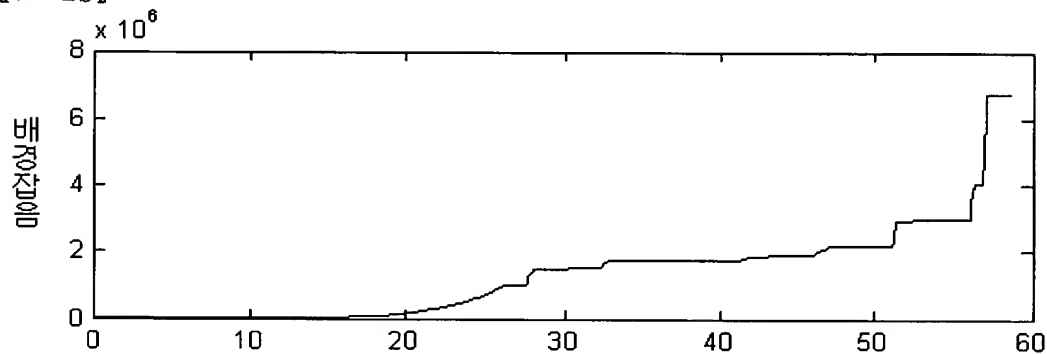
【도 1】



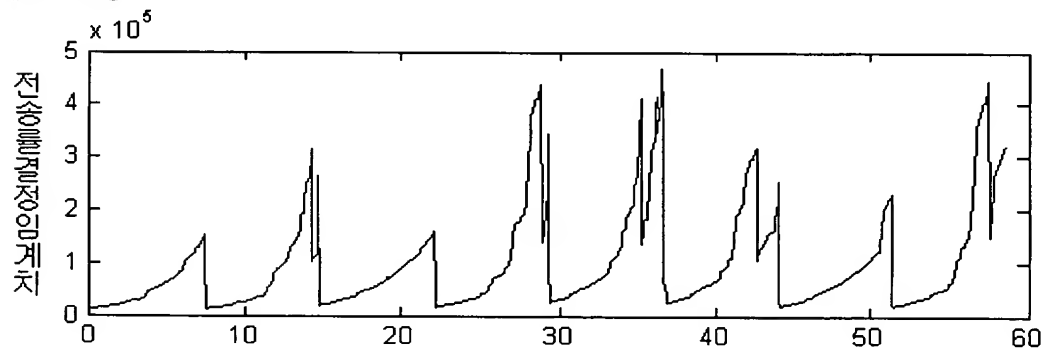
【도 2a】



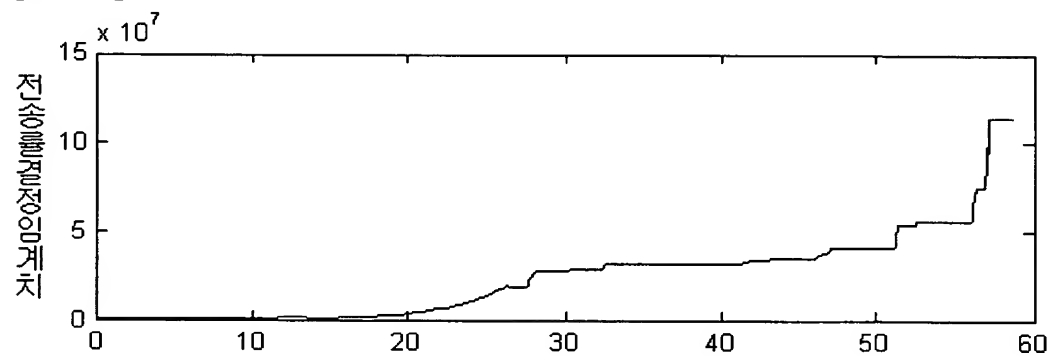
【도 2b】



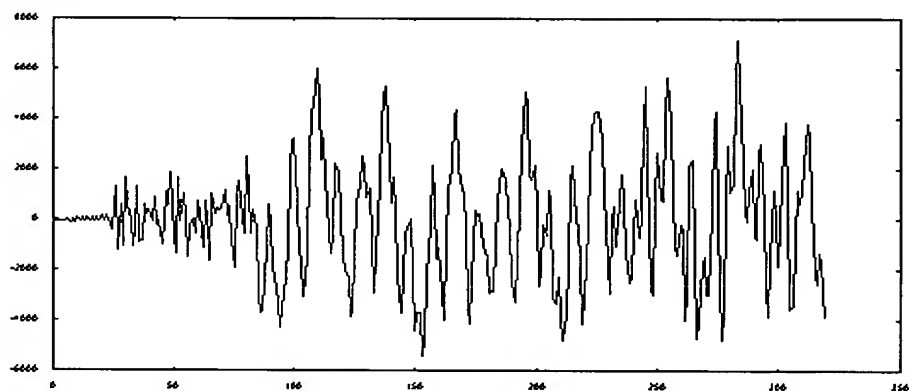
【도 3a】



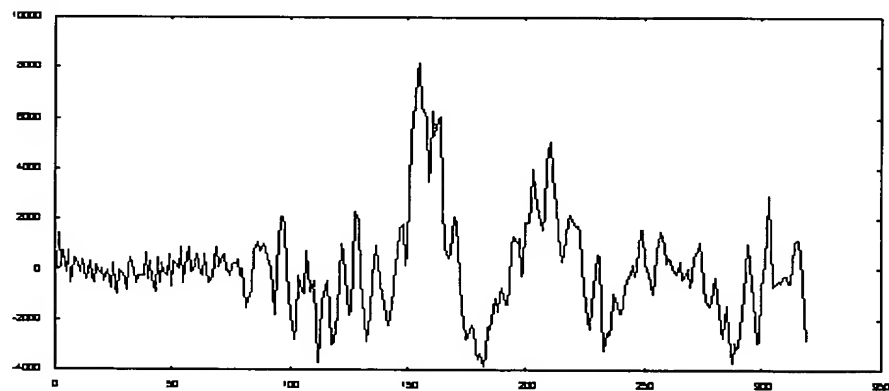
【도 3b】



【도 4a】

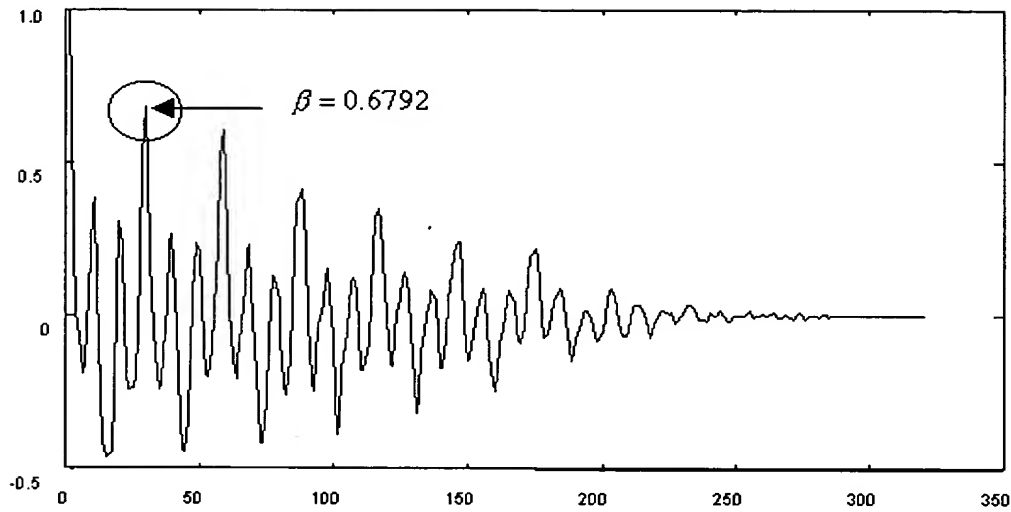


【도 4b】

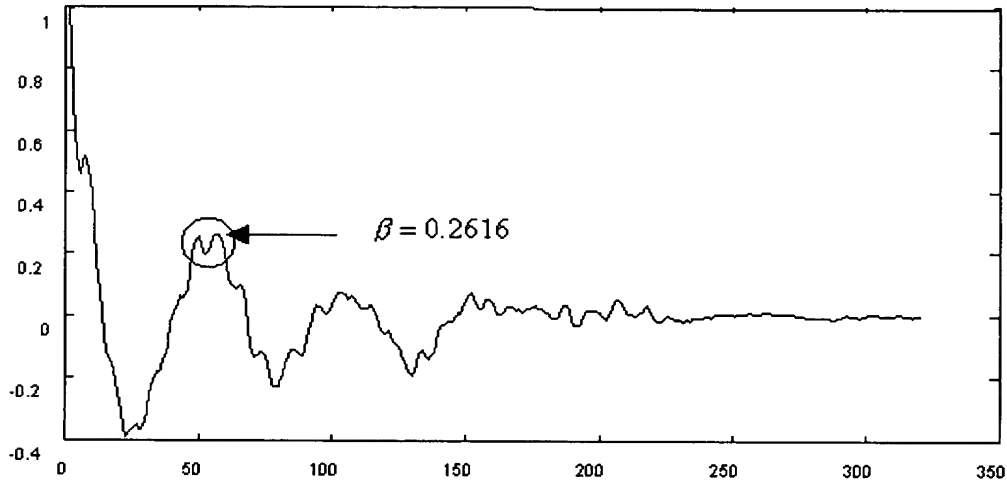




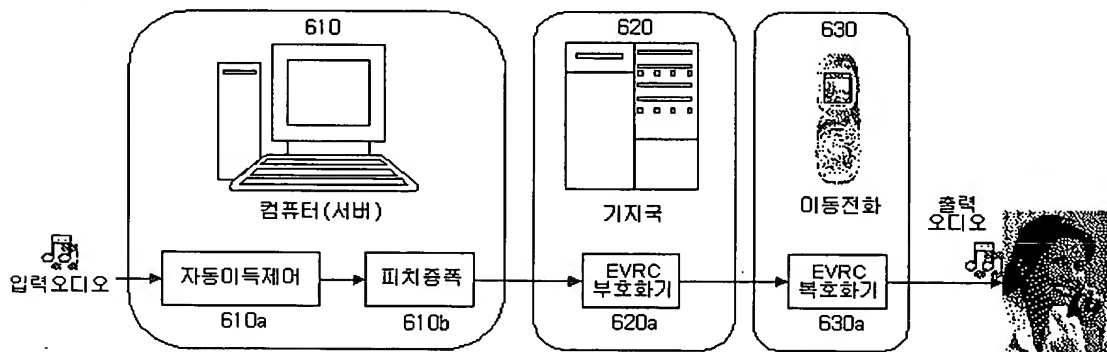
【도 5a】



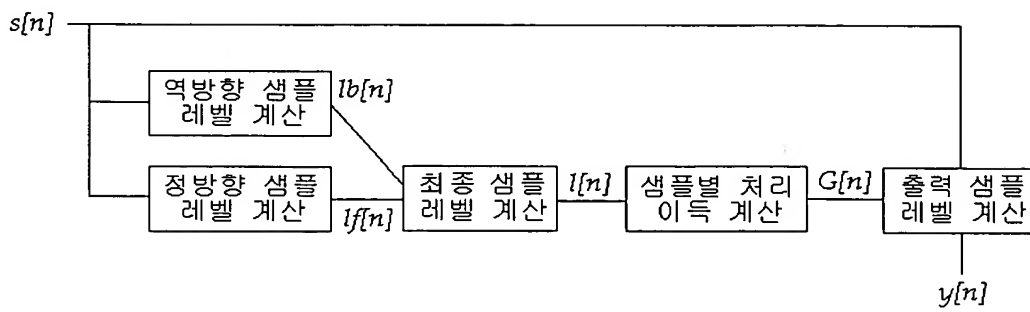
【도 5b】



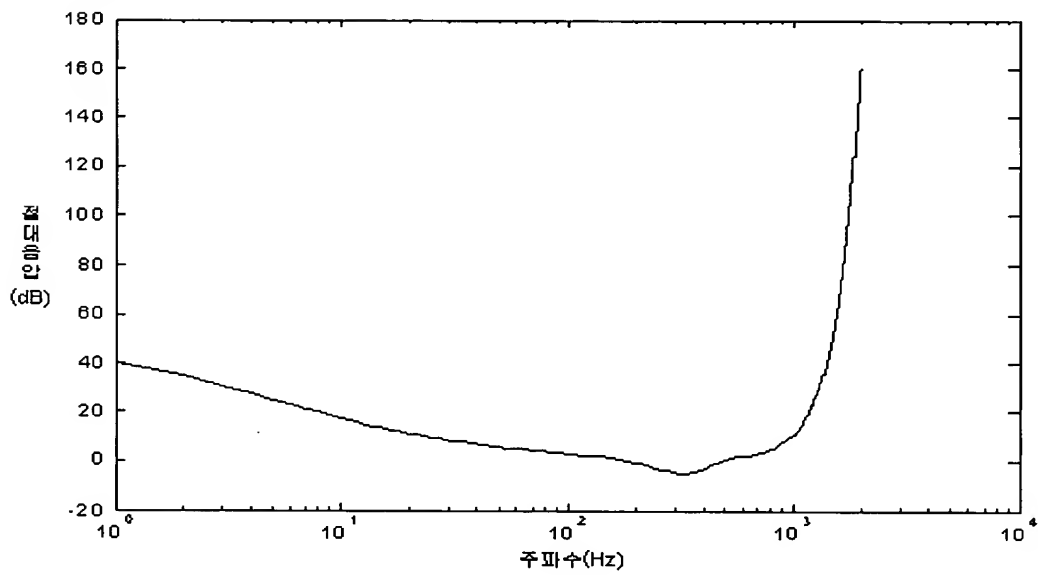
【도 6】



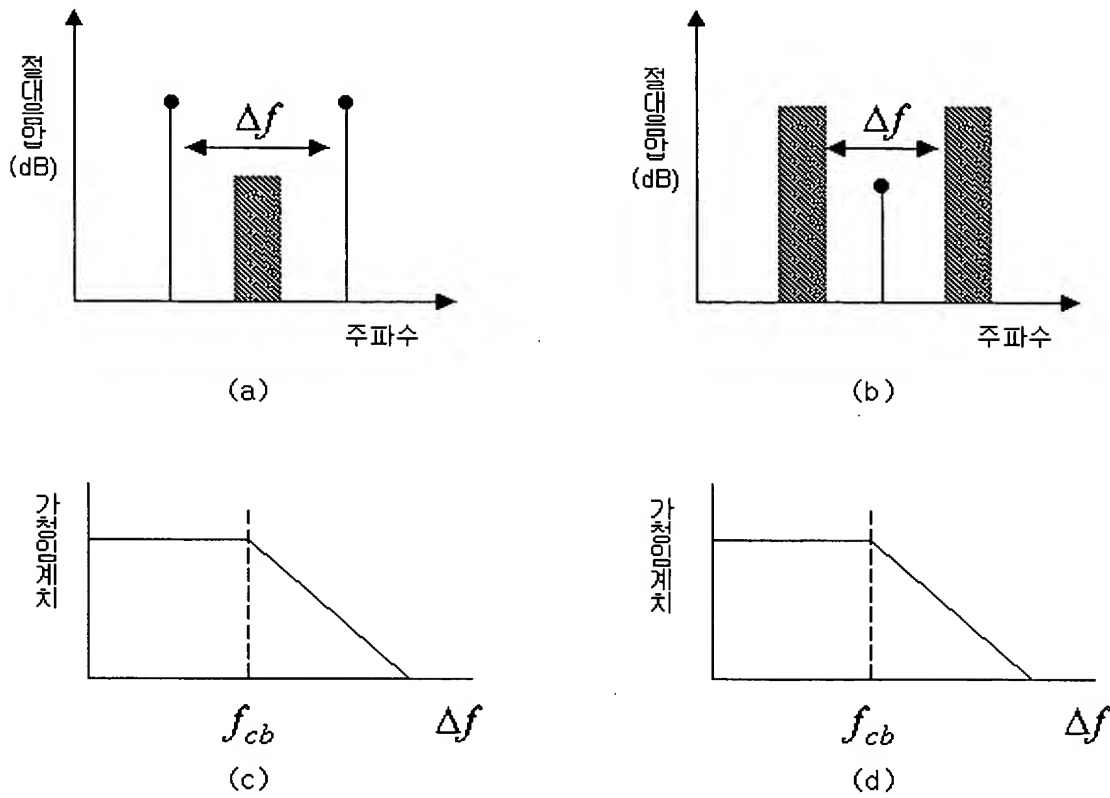
【도 7】



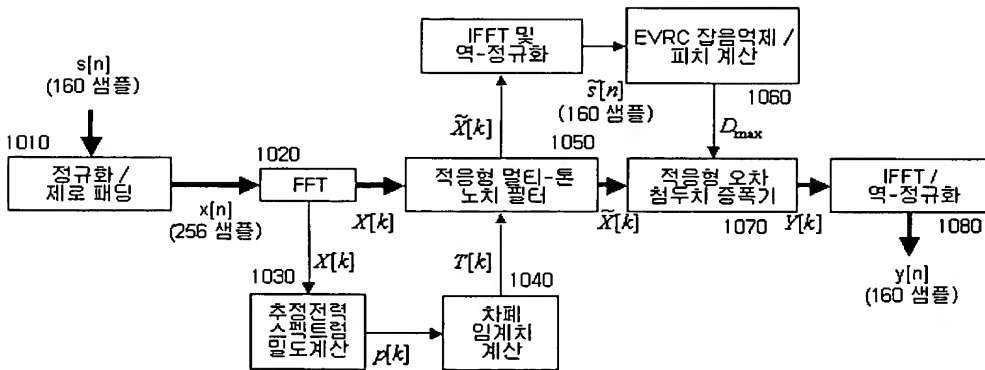
【도 8】



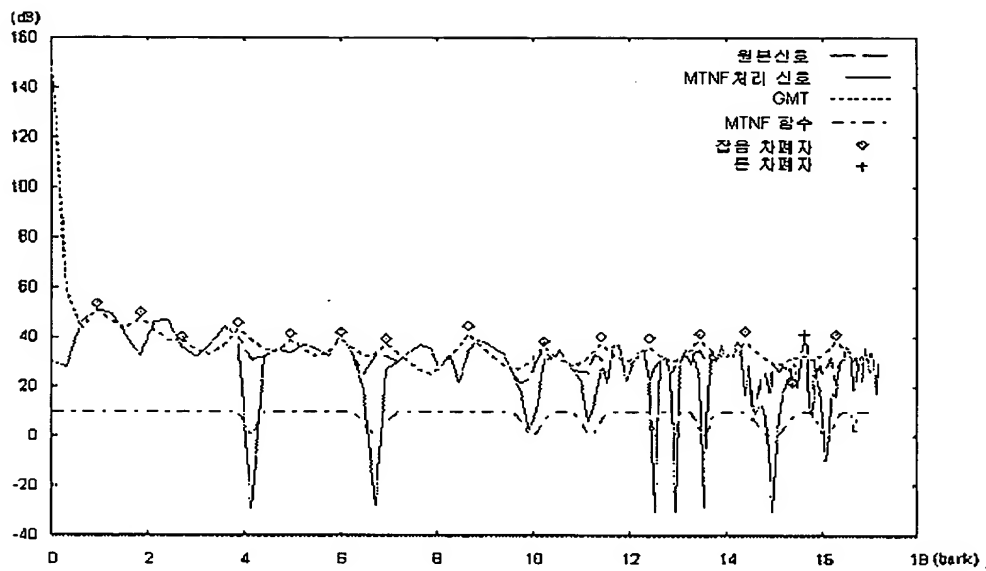
【도 9】



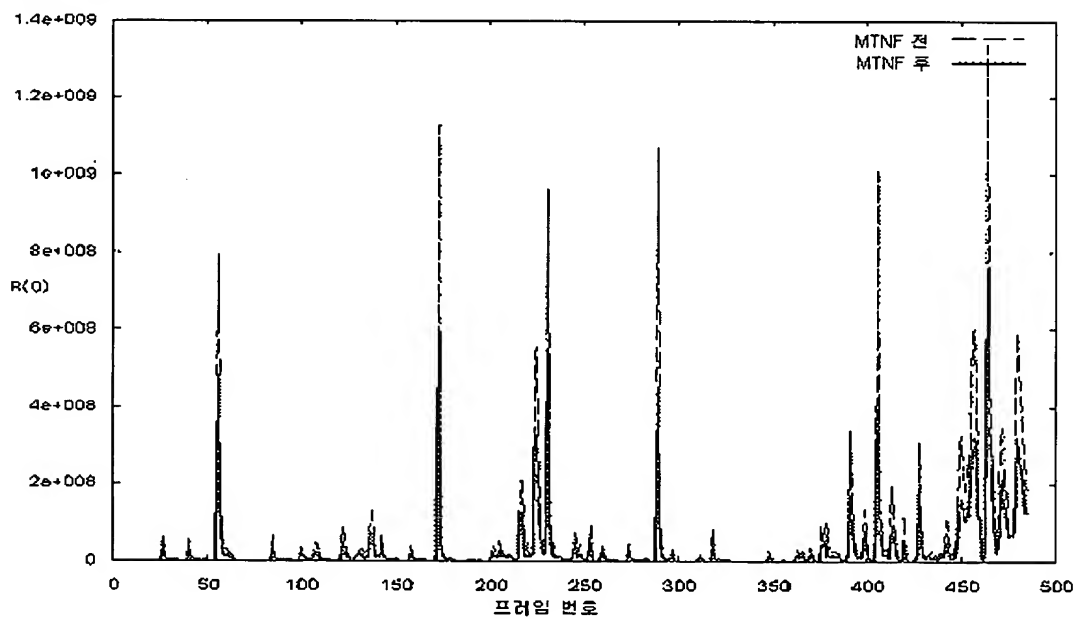
【도 10】



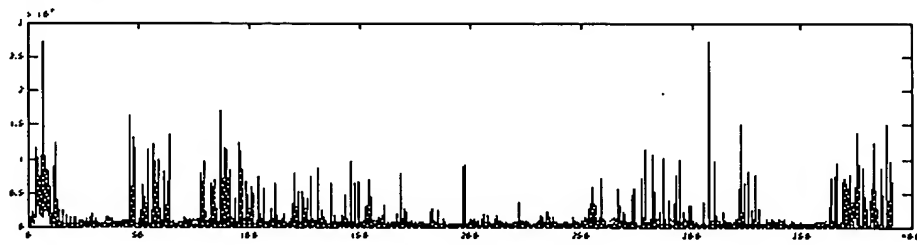
【도 11】



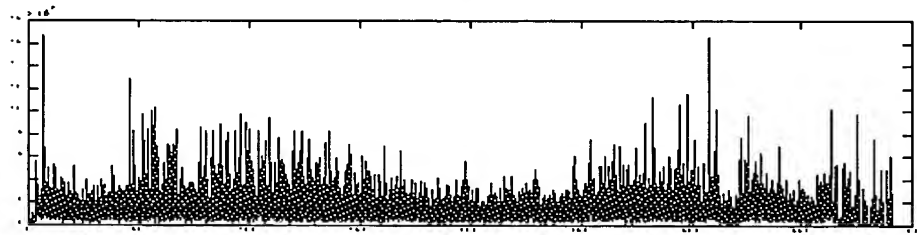
【도 12】



【도 13】



(a)



(b)

【도 14】

